面向商品评论文本的情感分析与挖掘[1]

李涵昱，周鹏飞

（中国科学院文献情报中心 信息系统部　北京　100190）

摘要　　随着电子商务的快速发展，互联网上出现大量商品评论信息，商品评论文本的情感分析与挖掘对于研究商品口碑、进行商品推荐都具有重要的价值。文中设计商品属性提取与过滤算法、情感词判别算法，分析商品的评论信息并自动抽取用户关注的商品属性和用户对相应属性的评价观点，进一步将其应用于商品评价文本的情感倾向性分析。在真实数据集上进行测试取得了准确率 81.08%，召回率 88.23%。

关键词　　商品评论挖掘、情感倾向性分析、情感词极性判断、数据挖掘

中图法分类号 TP391

Sentiment Analysis and Mining of Product Reviews

HanYu Li　PengFei Zhou

(National Science Library, Chinese Academy of Sciences , Information System Division ,BeiJing 100190)

Abstract　With the rapid development of e-commerce, a large number of product reviews arise on the Internet, and the sentiment orientation analysis and mining of product reviews become important for the study of the product reputation and recommending. In this paper, we design the algorithm of product attribute extraction and the word sentiment recognition. It will extract the product reviews and the corresponding evaluation point according to the analysis of the product reviews. Finally ,the sentiment orientation analysis of the product reviews is finished by the use of the above result. The system achieves the accuracy of 81.08% and the recall of 88.23% on the real data set and we give the result analysis.

Keywords　Mining of Product Review , Sentiment Analysis, Word Sentiment recognition, Data mining

---

[1]本文受基金项目 1 国家自然科学基金资助项目(4123091) ，
李涵昱（1986 年-），女，河南周口，硕士，助理馆员，主要研究方向为信息检索、自然语言处理（lihy@mail.las.ac.cn），周鹏飞（1985 年-），男，河北唐山，硕士，高级工程师，主要研究方向为数据挖掘、情感倾向性分析（hb1gzpf@126.com）。

1 引言

随着电子商务的快速发展，网络上产生了越来越多的商品评论数据。这些评论信息大多包含用户的主观情感，蕴藏着大量的价值等待人们挖掘。自动化的评论文本挖掘工具可以节省人工分析评论数据的时间，提高效率有很强的实用性[1][2]。面向商品评论文本的情感分析与挖掘目标是从文本中挖掘用户表达的观点和情感，抽取出用户的关注热点，有很大的实用价值[3]。生产厂家可以根据用户的评论信息发现产品的不足和消费者的需求，进而改进产品，销售商家可以根据用户评论来发现优质商品和用户的喜好，进而向用户推荐优质商品来提高销量，同时可以将抽取到的评价热点应用到商品搜索排序，用户可以根据汇总的评论信息发现产品的优势和不足找到自己最心仪的产品。

当前大部分文本情感倾向性分析不考虑文本的主题，篇章级情感倾向性分析主要为分析篇章级褒贬倾向[4][5]，词汇级情感倾向性分析主要研究一个词的褒贬倾向。篇章级情感倾向性分析主要有以下方法：（1）基于情感词典的方法，使用已有的情感词典，根据篇章中正负面情感词出现的数目决定整体情感倾向[6]；（2）基于机器学习方法如支持向量机、朴素贝叶斯、最大熵、K 近邻算法等[4]。文献[4]将情感倾向性分析作为文本分类来处理,通过使用标注好的训练数据训练分类器，然后使用分类器预测新的样本，取得了较好的效果。词汇级情感倾向性研究主要工作包括情感词典的构建、情感词极性的确定等[7][8][9]，文献[10]从人的角度出发用态度、影响、观点等方面对情感词进行更详细的划分。这些工作大多为主题无关的，在很多情况下无法满足用户需求，例如当用户希望得到消费者对商品的哪些功能最不满意的时候，篇章级情感分析是无法给出结果的。

当情感词在不同语境存在主题相关的情感倾向性分析，需要确定评价对象[11]，评价对象是指某段评论中所讨论的主题[12],具体表现为评论文本中评价词语所修饰的对象,如新闻评论中的某个事件或者产品评论中某种产品的属性(如“屏幕”)等。对商品评论的挖掘包括商品属性抽取、评论情感词提取等，通过提取这些知识可以进一步支持商品评价文本的情感倾向性分析[7]。目前相关工作研究对象主要为英文商品评论信息，使用的方法包括专家抽取、基于词性标注与词频统计、半监督方法等[1][2][13][14][15][16]。文献[13]采用弱监督的学习方法进行产品特征抽取，使用产品特征作为种子集合，从这些种子出现的语句中抽取文本模式，然后利用抽取到得文本模式抽取新的产品特征。文献[1][2][14]针对英文评论信息通过词性标注过滤得到商品属性，在评论文本中的商品属性附近发现情感词并进一步对评论观点进行归纳分析。

本文提出一种自动抽取商品属性及对应情感词的方法，基于抽取到的商品属性和对应情感词进行商品评论的情感倾向性分析。本文第 2 部分介绍系统的总体结构，第 3 部分介绍商品属性和情感词抽取模块，第 4 部分介绍商品评论的情感倾向性分析模块，第 5 部分介绍实验和结果分析，最后进行了总结并提出进一步研究方向。

2 面向评论文本的情感分析与挖掘系统

本文通过对商品评论文本中商品属性和对应评价情感词抽取，得到商品属性和情感词的修饰关系，确定情感词修饰商品属性时的情感倾向，进一步用来进行情感倾向性分析，系统整体流程如图 1 所示。该方法无需种子属性集合，利用商品属性和情感词的共现关系得到商品属性和情感词，抽取到更加可靠的商品属性和情感词信息。本系统除输入商品评论文本外无任何领域相关知识，无需任何修改，只要提供新的领域评价数据就可以应用到其他领域，可移植性好。

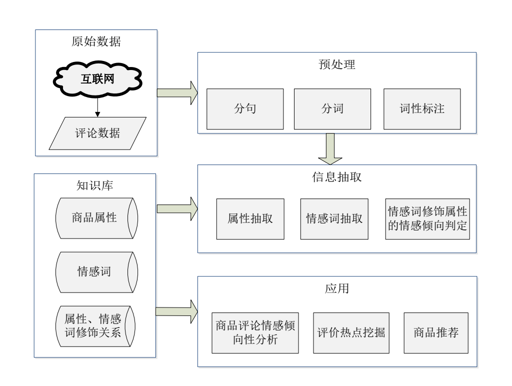


图 1 商品评论情感分析结构

系统输入是用户对某种商品的评论信息，例如购物网站的某款手机评论，包括用户的评论文本、用户对该评价给定的星级。系统通过预处理、分句、分词将用户评论信息转换为较短的评论短语，根据词性标注结果将其中名词作为候选商品情感词，将形容词作为候选情感词，根据商品属性和情感词共现原理抽取得到商品属性和每条属性对应的情感词，同时根据原始评论文本的星级来得到情感词的正负面 ，部分属性抽取和情感词抽取结果如图 2 所示，左边表示抽取到得商品属性如手机的属性包括“屏幕”、“电池”、“内存”、“价格”等，右边表示属性对应的情感词，例如“屏幕”属性抽取到得情感词包括“大”、“清晰”、“灵敏”、“垃圾”等。

通过抽取到得这些商品属性以及对应情感词可以为商品建立一个情感分析的知识库，这个知识库可以用来进行新的商品评论情感倾向性分析，建立评价属性与情感词的对应关系可以解决情感词修饰不同属性情感倾向不同的问题。同时，在此基础上进行深层挖掘，例如找出用户比较关注的商品属性、找到用户最不满意的商品属性、找到用户差评最多的商品属性等[15]。

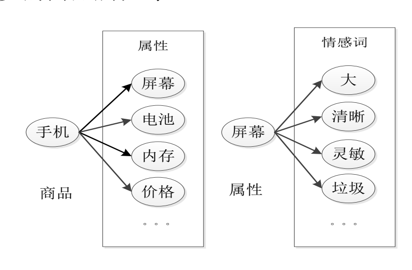


图 2 商品属性与情感词抽取结果示例

3 商品属性和情感词发现

大部分的商品评论数据比较简短，如“手机屏幕很大”，仅对一个属性进行评价，有的评论数据比较长，可能包含对多个属性的评价，我们将分析该类数据，进行商品属性和情感词的发现。本文使用 ICTCLAS 汉语分词系统进行分词和词性标注。评论数据中的商品属性主要为名词形式，例如“屏幕”、“电池”、“声音”，因此将词性标注得到的词语序列中的名词等作为候选的属性词。同理情感词一般为形容词，将抽取到的形容词等作为候选的情感词。根据名词出现的频率和与情感词语的共现等方法将噪音数据过滤，候选商品属性和情感词抽取算法如表 1 所列。

表 1 候选商品属性和情感词抽取算法

输入：评价短语集合 Segments = $\{seg_1, seg_2 \cdots seg_n\}$，α(出现频率的阈值)。
输出：商品属性集合 $F=\{f_1, f_2, \cdots. f_n\}$，每个商品对应评价情感词集合 $S=\{s_1, s_2, \cdots s_m\}$。
步骤：
1. 初始化候选商品属性集合 F 为空，候选情感词集合 S 为空。
2. 对于每个评价短语，抽取出名词短语，放入属性集合 F。
3. 对于每个评价短语，抽取出形容词短语，放入情感词集合 S。
4 . F 中 f_i 元素，若 f_i 在评价数据中出现频率低于 α，则从集合中删除。
5. S 中 s_j 元素，若 s_j 在评价数据中出现频率低于 α，则从集合中删除。
4. 从 F 和 S 中任意取一对元素 $\langle f_i, s_j \rangle$，如果这两个元素同时出现在某评价短语中，那么这两个元素组成一个评价对，所有评价对组合组成评价对集合 PairSet。
5. 对于 F 中所有元素，如果没有出现在 PairSet 中，从 F 中除去该元素。
6. 对于 S 中所有元素，如果没有出现在 PairSet 中，从 S 中除去该元素。
7. 经过如上过滤后，得到商品属性集合 F，情感词集合 S。

4 商品评论情感倾向性分析

4.1 评论数据基准情感词抽取与倾向判断

通过对评价数据分析，发现很多情感词如“不错”、“很好”、“很差”、“不好”等，这些词出现的频率很高，情感倾向明确，这里将该类情感词抽取出来作为基准情感词使用。由于抽取到的商品评价数据的情感倾向可以根据评价的星级确定，我们利用已有评价数据的情感倾向来确定情感词的情感倾向。使用开方检验（CHI）方法计算情感词和正负面类别的关联程度，开方估计值越大表示情感词的情感倾向越可靠。每一个商品属性的每一个情感词分别计算 CHI 值，在类别中情感词的 CHI 值如公式 1 计算。

$$\chi^2(t,c)=\frac{N\times(AD-BC)^2}{(A+C)(B+D)(A+B)(C+D)} \quad (公式1)$$

其中，N 为训练语料中的文档总数，c 为某一类别，t 为某情感词，A 为属于类别 c 且包含情感词 t 的文档个数，B 为不属于类别 c 但包含情感词 t 的文档个数，C 为属于类别 c 但不包含情感词 t 的文档个数，D 为不属于类别 c 且包含情感词 t 的文档个数。分别计算情感词 t 和正负两个类别的关联度，关联度较大的类别作为该情感词 t 的情感倾向。

对于每个商品属性抽取出对应的情感词，并对情感词按照出现频率进行排序，将出现频率与 CHI 值大于阈值的情感词抽取出来作为基准情感词，这里使用的频率阈值为 0.1，CHI 阈值为 0.9，基准情感词的抽取更加注重准确率利用抽取到的基准情感词集进一步扩展将提高召回率。

4.2 评价短语的情感倾向判断

利用抽取到的基准情感词集进一步扩展情感词集合规模，计算情感词和基准情感词的距离主要使用 PMI-IR 算法等[17]。PMI-IR 算法计算公式如公式 2 所示。

$$PMI(word_1,word_2)=\log(\frac{P(word_1\ \&\ word_2)}{P(word_1)P(word_2)}) \quad (公式2)$$

其中 P($word_1$&$word_2$) 表示 $word_1$ 和 $word_2$ 同时出现的概率，P($word_1$) 表示 $word_1$ 出现的概率。

当计算一个情感词的情感倾向时可以计算该词和基准情感词表中正负面情感词表中所有词的距离[17]，如公式（3）所示。

$$SO-PMI(word)=\sum_{pword\in Pset}PMI(word,pword)-\sum_{nword\in Nset}PMI(word,nword) \quad (公式3)$$

其中 word 为待计算情感倾向的新词，Pset 为基准情感词的褒义情感词集合，Nset 为基准情感词的贬义情感词集。

使用这种方法可以计算特定领域中一些情感倾向很明确的情感词的情感倾向，然而同一个情感词修饰不同商品属性可能会出现情感倾向相反的情况，例如：

“手机屏幕很大”　“手机噪音很大”

这里“大”同时作为“屏幕”和“噪音”两个属性的情感词。当作为修饰“屏幕”的情感词时为正面情感词，当作为修饰“噪音”的情感词时为负面情感词。如果将“大”作为一个情感词计算与基准词的距离，无论计算结果如何都不准确。这里对 PMI 方法进行改进，不再仅计算情感词之间的距离，而是计算<属性、情感词>Pair 与基准情感词之间的距离来得到<属性、情感词>Pair 内情感词的情感倾向，上例中不再计算“大”与基准情感词的 PMI 值而是计算出现<屏幕，大>这个评价短语与基准情感词集的 PMI 值，改进的方法如公式（4）。

$$SO-PMI(feature,word)=\sum_{pword\in Pset}PMI(<feature,word>,pword)-\sum_{nword\in Nset}PMI(<feature,word>,nword) \quad (公式4)$$

计算 word 作为情感词修饰 feature 属性时，word 的情感倾向， Pset、Nset 含义同公式（3）。

4.3 商品评论情感倾向性分析

抽取到的商品属性和每个属性对应的情感词可以作为进行情感倾向性分析的知识库进行商品评论的情感倾向性分析[18]。例如通过对手机评论信息进行抽取可以得到手机的商品属性、每条属性对应的评价情感词语和使用这些情感词评价属性时的情感倾向性。当遇到相同领域的一条评论文本时处理方法如图 3 所示。

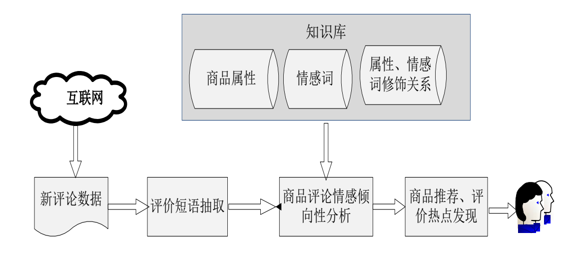


图 3 商品评论情感倾向性分析

对于每条评价短语，基于知识库抽取出评价属性。例如知识库中已经有抽取到的“屏幕”、“电池”、“系统”等属性，而评价短语中出现了“屏幕”这个属性词，则将“屏幕”作为抽取到的属性词。

对于每条评价短语，基于知识库抽取出评价情感词。根据知识库中商品属性和评价情感词组合的情感倾向得出评价短语的情感倾向。将每条评价短语的情感倾向汇总得出整体评价倾向。

商品属性评论观点汇总，由提取到的商品属性和属性对应的评价情感词可以汇总得到用户对于某一个商品的评价，如价格对应的评价有“实惠”、“便宜”、“贵”、“合适”、“合理”、“不错”等，通过发现的商品属性和对应的情感词可以得到每个属性对应的评价观点，分析各种评价观点的出现频率可以发现用户对于属性的评价热点。

5 实验与分析

5.1 实验数据

本文实验数据通过抓取电商网站京东商城(www.jd.com)一些比较热门的关于手机商品评论数据。网站上的用户评论数据包含对商品的评论文本信息和用户对商品评价的星级，用户对商品评价越高则给出的星级越高，反之，星级越低，星级从 1（低）到 5（高）分布如表 2 所示。

表 2 实验数据评论星级分布

商品	一星	二星	三星	四星	五星
商品 1	2	2	6	25	65
商品 2	3	1	8	31	57
商品 3	1	0	3	23	73

5.2 评价指标

本文采用准确率、召回率及 F 值作为评价指标。

准确率：

$$P = \frac{tp}{tp + fp} \qquad \text{（公式 5）}$$

其中 tp 为分析结果中正确的数目，fp 为分析结果中错误的数目。例如评价商品属性抽取的准确率时，tp 表示抽取到的商品属性中正确的数目，fp 为系统抽取结果认为是商品属性但实际上不是商品属性的数目。

召回率：

$$R = \frac{tp}{tp + fn} \qquad \text{（公式 6）}$$

其中 tp 为分析结果中正确的数目，fn 为正确的但没有被识别出的数目。例如评价商品属性抽取召回率时，tp 表示抽取到的商品属性中确实为商品属性的数目，fn 为在评论中出现但没有抽取到的商品属性数目。

调和平均值：

$$F = \frac{2}{\frac{1}{P} + \frac{1}{R}} = \frac{2PR}{P + R} \qquad \text{（公式 7）}$$

5.3 商品属性与评价短语抽取结果分析

实验数据中发现用户关注属性主要集中在“质量”、“屏幕”、“电池”、“价格”等方面。实验抽取到的部分商品属性和评价短语如表 3 所示。

表 3 抽取属性与评价短语示例

属性	正面评价短语	负面评价短语
屏幕	屏幕大、屏幕清晰、屏幕细腻、屏幕不错、屏幕灵敏…	屏幕小、屏幕暗、屏幕一般…
质量	质量很好、质量放心、质量不错、质量可靠…	质量太差、质量不好、质量不行…
电池	电池耐用、电池够用、电池给力、电池不错…	电池发热、电池不好、电池不行、电池很差、电池不经用…
价格	价格便宜、价格实惠、价格划算、价格公道…	价格贵、价格偏高…
系统	系统好用	系统老、系统慢…

三件商品，每件商品 100 条商品评价数据，抽取商品属性经过人工评价，使用表（1）算法中不同的 α 阈值采用召回率、准确率、调和平均值表示实验结果，得出结果如图 4 所示。三件商品，每件商品 100 条商品评价数据，抽取评价短语经过人工评价，使用表（1）算法中不同的 α 阈值采用召回率、准确率、调和平均值表示实验结果，得出结果如图 5 所示。

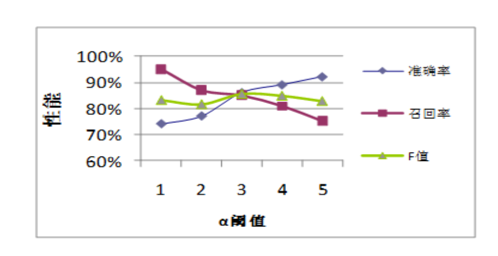


图 4 商品属性抽取结果评价

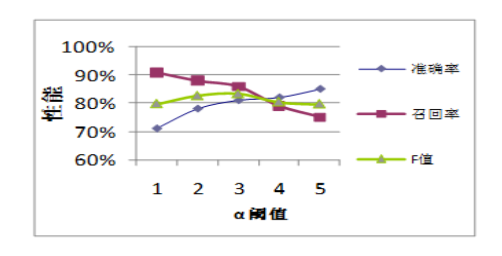


图 5 商品评价短语抽取结果评价

从图 4、图 5 中可以看出当阈值 α 较小时，较多冗余数据存在准确率较低，召回率较高，当阈值 α 较大时，能冗余数据减少，但是一些正确的结果也被过滤掉，召回率较低，准确率较高，在阈值为 3 时取得较好结果。通过观察可以发现召回率比较高，因为某些很多评论中出现了相同的属性信息，虽然在某些评论中的属性没有抽取出来，但是在其他评论中抽取出来之后可以相互弥补，所以整体召回率较高。准确率比召回率要低一些，通过进一步改进算法还可以进一步提高准确率。

5.4 商品评论倾向性分析

抽取到商品属性和每个属性对应的情感词以后就可以使用抽取到的信息进行情感倾向性分析，对于每条评论数据分析的结果为：这条评论信息评价的商品属性以及对该属性的情感倾向是正面还是负面。这里使用 1700 篇商品评论数据抽取商品属性、情感词、确定情感词修饰商品属性时的情感倾向，然后使用这个知识库进行情感倾向性分析，对三种商品的评论数据进行实验，每个商品有 100 条评论进行情感倾向性分析，人工对分析结果进行评价结果如表 4 所列。

表 4 商品评论情感倾向性分析结果表

商品名	召回率	准确率	*F 值*
手机商品 1	85.18%	79.31%	*82.14%*
手机商品 2	85.71%	80.00%	*82.76%*
手机商品 3	*88.23%*	*81.08%*	*84.50%*

从实验结果可以看到，不同商品的倾向性分析结果，F 值均在 80%以上，表明使用这种方法进行商品评论的情感倾向性分析切实可行。进行情感倾向性分析之后，对不同的评价观点按照各个商品属性进行分析，如图 6、图 7、图 8 所示。

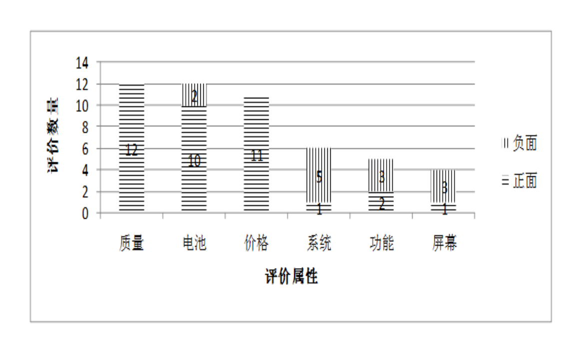


图 6 手机商品 1 评价情感倾向性分析结果

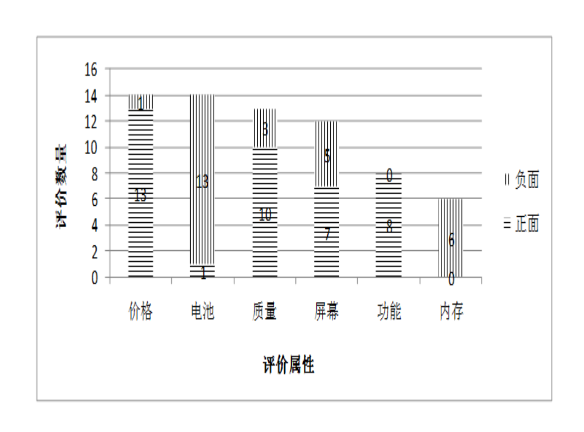


图 7 手机商品 2 评价情感倾向性分析结果

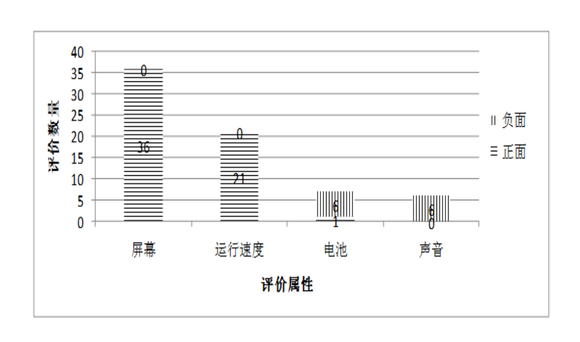


图 8 手机商品 3 评价情感倾向性分析结果

手机 1 为一款诺基亚传统功能手机，从上面可以看出用户的评价热点反应出了该商品的特点：手机质量很好，价格便宜，电池够用，屏幕较小、功能较少、系统较旧。用户对质量、价格非常满意，对电池比较满意，对屏幕、功能、系统不太满意。手机 2 为一款入门智能机，从分析结果可以看出用户对价格和功能很满意，对质量和屏幕基本满意，对电池和内存不满意，这样的分析结果基本反映了该手机的特点。手机 3 为一款大屏智能机，从分析结果可以看出用户反映最多的是屏幕很好、运行速度很满意，用户对这两个方面的反馈最多，同时用户指出电池和手机声音不好。分析结果较好地反映了各款手机的相对特点，对用户、商家、生产厂家有很好的参考价值。

6　结论

本文实现了自动化的商品属性和评价情感词抽取，实现了商品评论的情感倾向性分析，性能达到 80%以上，实验结果表明该方法取得较高准确率和召回率，然而有些方面还需要进一步完善改进，某些评价属性和评价观点是类似的需要进一步聚类，例如“价格合适”与“价格合理”应该归并为一条。通过本文工作完成了对商品评论中关于商品属性的评论挖掘，抽取出了评论的属性并判断了评论的正负面，后续工作可以根据情感倾向性分析结果向用户推荐优质商品等应用。

参考文献

[1] Hu Minqing, Liu Bing. Mining Opinion Features in Customer Reviews. Proceedings of the National Conference on Artificial Intelligence,2004[C].San Jose, USA: ACM Press, 2004: 755-760.

[2] Bing Liu, Minqing Hu, Junsheng Cheng. Opinion Observer: Analyzing and Comparing Opinions on the Web. Proceedings of the 14th international conference on World Wide Web, 2005[C].ACM,2005:342-351 .

[3] Lisette Garc´ıa-Moya, Henry Anaya-S´anchez,Rafael Berlanga-Llavori.Combining Probabilistic Language Models for Aspect-Based Sentiment Retrieval. Proceedings of the 34th Eu-ropean Conference on Information Retrieval, 2012[C].2012:561-564.

[4] B. Pang, L.L. Lee and S. Vaithyanathan. Thumbs up? Sentiment Classification Using Machine Learning Techniques. Proceedings of ACL-02 conference on Empirical methods in Natural Language Processing, 2002[C].Volume 10, 2002:79-86.

[5] Bo Pang, Lillian Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. Proceedings of American Association for Artificial Intelligence Conference, 2004[C],2004:271－278.

[6] 张成功，刘培玉，朱振方，方明。一种基于情感词典的情感分析方法[J]。山东大学学报,2012，第 47 卷，第 3 期。

[7] Yu, J, Zha, Z, Wang, M, Chua, T.Aspect ranking: Identifying important product aspects from online consumer reviews.[J] Proceeding of Computational Linguistics, 2011, 1496–1505 (2011).

[8] Vasileios Hatzivassiloglou, Kathleen R.McKeown. Predicting the Semantic Orientation of Adjectives. Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics,1997[C].1997:174-181.

[9] 朱嫣岚，闵锦，周雅倩，黄萱菁，吴立德。基于HowNet 的词汇语义倾向计算[J]。中文信息学报,2006, 20(1): 14-20.

[10] Casey Whitelaw, Navendu Garg, Shlomo Argamon. Using Appraisal Groups for Sentiment Analysis. In Proceeding CIKM '05 Proceedings of the 14th Association for Computing Machinery international conference on Information and knowledge management, 2005 [C]. 2005: 625-631.

[11] 赵妍妍，秦兵，刘挺。文本情感分析[J]。软件学报，2010,21（8）：1834-1848.

[12] 姚天昉，程希文，徐飞玉，汉思．乌思克尔特，王睿。文本意见挖掘综述[J]。中文信息学报，2008, 28(11): 2725-2728。

[13] 伍星，何中市，黄永文。基于弱监督学习的产品特征抽取[J]。计算机工程，2009,13：199-201.

[14] Hu Minqing, Liu Bing. Mining and summarizing customer reviews. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining,2004[C]. New York, NY, USA 2004: 168-177.

[15] Lee, D., Jeong, O., Lee, S. Opinion Mining of Customer Feedback Data on the Web. Proceedings of the 2nd international conference on Ubiquitous information management and communication,2008[C]. ACM Press New York,2008: 230-235.

[16] Dave, K., Lawrence, S., Pennock, D. M. 2003. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. Proceedings of the 12th international conference on World Wide Web,2003[C].ACM Press New York, 2003: 519-528.

[17] 张猛，彭一凡，樊扬，李丹，林小俊，吴玺宏。中文倾向性分析的研究。第一届中文倾向性分析评测研讨会，[C],2008，38-45.

[18] Yue Lu , Malu Castellanos, Umeshwar Dayal , ChengXiang Zhai. Automatic Construction of a Context-Aware Sentiment Lexicon: An Optimization Approach. Proceeding of the 20th international conference on World Wide Web,2011[C]. 2011:347-356.

作者联系方式：李涵昱 中国科学院文献情报中心信息系统部 15201227359 lihy@mail.las.ac.cn